DOCUMENT RESUME

ED 406 448 TM 026 428

AUTHOR Dean, Linda M.
TITLE An Evaluation Criteria Validation Model: Stakeholder Participation in Evaluating a Set of Criteria.
PUB DATE [97]
NOTE 14p.; Paper presented at the Annual Conference of the American Evaluation Association (Atlanta, GA, November 6-9, 1996).
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Criteria; Data Analysis; Data Collection; *Evaluation Methods; Evaluation Utilization; Models; Program Evaluation; Test Construction; *Validity
IDENTIFIERS *Stakeholders

ABSTRACT

This paper proposes a model to establish which criteria are considered by stakeholders as valid for evaluating a program. The model is developed with the aim of increasing the credibility and use of evaluations. Stakeholders are involved in the identification of potential evaluation criteria, and ratings of validity and priority are used as the basis for criteria selection. The Evaluation Criteria Model is composed of the following six steps: (1) identify stakeholders; (2) identify potential evaluation criteria; (3) develop survey instruments; (4) collect data; (5) analyze data; and (6) select set of criteria. The first base of the criteria validation model is the job analysis process used in business and industry to document the content-related validity of personnel evaluation tests. The second base is the front-end activity of validating program evaluation issues and concerns that was proposed by E. Guba and Y. S. Lincoln (1981). These authors suggest a formal survey of stakeholders, including ratings of issues and concerns, to identify the areas to address. These approaches are combined in the proposed model. (Contains three figures and seven references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
LINDA M. DEAN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

AN EVALUATION CRITERIA VALIDATION MODEL:
Stakeholder Participation in Validating a Set of Criteria

Linda M. Dean

This paper proposes a model to establish which criteria are considered by stakeholders as valid for evaluating a program. The model is developed with the aim of increasing the credibility and use of evaluations. Stakeholders are involved in the identification of potential evaluation criteria and ratings of validity and priority are used as the basis for selection of a set of evaluation criterion.

ED 406 448

Evaluation literature is replete with references to the importance of involving program stakeholders in program evaluation activities. Stakeholders may be included in such evaluation activities as determining what information is desired (Stake, 1972), validating the evaluation issues to be addressed (Guba & Lincoln, 1981), or as a data source for outcomes attained.

Numerous authors have also suggested that evaluation criteria need to be considered as valid in order for stakeholders to use an evaluation. Michael Scriven (1974) addressed the importance of selecting evaluation criteria and the complication of multiple values over twenty years ago. There is evidence that an evaluation will be perceived as credible to stakeholders if the criteria are perceived as valid (Greene, 1985).

Even though we acknowledge the importance of involving stakeholders and the importance of selecting evaluation criteria that are perceived as valid by stakeholders, there is little reference in the literature to how this might be achieved. The purpose of this paper is to propose an Evaluation Criteria Validation Model which evaluators may use to select a set of criterion. The process begins with selection of stakeholders and concludes with selected criteria and documented evidence of which criteria the stakeholders consider as valid for the evaluation of the specified program.

TM026428

BEST COPY AVAILABLE

The Evaluation Criteria Validation Model is comprised of six steps:

1. Identify Stakeholders
2. Identify Potential Evaluation Criteria
3. Develop Survey Instrument(s)
4. Collect Data
5. Analyze Data
6. Select Set of Criterion

A graphic over-view is portrayed in an attachment. The model draws upon literature from three sources; program evaluation, educational testing and business.

Bases for the Model

Two specific validation processes serve as the bases for the proposed criteria validation model. The first base is the job analysis process which is used in business and industry for documenting the content-related validity of personnel evaluation tests. Job analysis is also used in professional licensure testing for establishing the licensure examination content specifications. This methodology has been upheld in courts of law as appropriate for establishing the content-related validity of licensure examinations (Thompson & Thompson, 1982). Several details of this process are included later in this paper.

The second base for the proposed model is the front-end activity of validating program evaluation issues and concerns that was proposed by Guba and Lincoln in Effective Evaluation (1981, p.319-320). The authors suggest a formal survey of stakeholders, including ratings of issues and concerns, in order to identify which issues and concerns will be addressed. More details of this process are included later in this paper.

Validating Examination Content

As indicated, in large scale licensure testing, ratings from a job analysis serve as the basis for establishing examination content-related validity (Byham & Spitzer, 1971). A representative

sample of practicing professionals completes a questionnaire that lists the tasks required for the job and asks for two or three ratings, such as, frequency of performing the task, importance of the task, criticality if the task is performed improperly or amount of time spent. Tasks with ratings below a pre-established point are eliminated from further consideration. Ratings can be combined into an "index of importance" (Kane et all, 1989). Examination content specifications are then weighted to reflect the findings of the job analysis.

Validating Evaluation Issues or Concerns

The front-end activity of validating program evaluation issues and concerns was proposed by Guba and Lincoln (1981, p.319-320). In this activity, a large representative sample of stakeholders is asked to complete a questionnaire that lists each issue or concern and asks for two ratings; first, the degree of validity; and, second, the priority. Items that receive a low rating on validity or priority from any audience are eliminated from further consideration. The authors suggest that items with a large variance in validity ratings deserve further study. The authors caution not to determine priority simply by aggregation of ratings over all audiences as variance in priority may reflect values. The survey of stakeholders is an integral part of the proposed validation model.

Validating Evaluation Criteria

The proposed model includes identifying the potential criteria, preparing an instrument, obtaining ratings of validity and priority from a representative sample, analyzing the data and combining the data into a framework for interpretation. This paper operationalizes the program evaluation criteria validating activity. Each step is broken out into tasks that are identified with a corresponding number and sequential letter. A detailed description of each step and task follows.

Model Step 1: Identify Stakeholders

The purpose of Step 1 is to identify a base of individuals who could serve as informants for identifying potential evaluation criteria (Step 2) and/or respondents to the survey questionnaire (Step 4). There are three tasks in Step 1.

Task 1a. Identify pertinent stakeholder groups -- The evaluator identifies pertinent stakeholder groups by reviewing documents and conducting a series of brief, informal inquiry interviews with such contact persons as the client, administration personnel, faculty, and other such persons identified through the review of documents. If there are a large number of pertinent stakeholder groups, a categorization scheme may be appropriate.

Task 1b. Build stakeholder data base -- Locate mailing lists, telephone lists, group membership lists and other data bases. Inquiry is also made regarding the availability of mailing labels or distribution resources that can be used in either Step 2 or Step 4.

Task 1c. Select stakeholder informants -- A small sample of informants is selected with the intent of optimizing information rather than being a representative sample (Guba & Lincoln, 1981, p.319). The selected informants need to be directly familiar with the program and able to articulate potential evaluation criteria that may be important to each stakeholder group (Step 2). Personal contact is made with prospective informants to gain information about their experience in and depth of knowledge about the subject program. A brief description of the validation process is offered and a preliminary reading regarding potential participation is made. From this information, the evaluator selects a small group of stakeholder informants, hereafter referred to as informant group. Obtaining an agreement to serve and a commitment to the process concludes this task.

Model Step 2: Identify Potential Evaluation Criteria

The purpose of Step 2 is to develop a list of potential evaluation criteria that will be used in the survey instrument(s). The list needs to be sufficiently inclusive in order to capture those criteria that may be deemed important by members of the various stakeholder groups. Step 2 is comprised of three tasks.

Task 2a. Affirm the primary evaluation question(s) -- Review documents and interview pertinent stakeholders to affirm the primary evaluation question(s). Depending upon the circumstances of the evaluation, there may be multiple purposes and multiple evaluation questions. When the evaluation is attempting to answer multiple questions, the survey instrument (Step 3) may be designed to accommodate more than one question. As will become evident, this process accommodates a limited number of evaluation questions.

Task 2b. Collate initial list of potential criteria -- The evaluator initiates a collated list of potential criteria by using such data collection methods as a review of documents, interviews, informal observation, review of literature and by drawing upon this own personal evaluation experience. If a large number of potential criteria are identified, say 15-30 criteria, a categorization scheme may be required. Because the purpose of this step is to identify many viable potential criteria, categories of criteria may be suggested to the interviewee if the interviewee expresses hesitation or uncertainty. Follow-up questions can be formed as the discussion progresses, in order to clarify and specify fully the evaluation criteria. The degree of specificity is an integral part of this step in order that each potential criterion be measurable.

Task 2c. Finalize potential criteria list -- Once formulated, the list is then reviewed by the informant group. The goal is to review the criteria list for clarity, specificity and completeness. If resources are limited, a review by a few key individuals, may suffice.

It is noted that accompanying standards are required if the evaluation is to provide conclusions or recommendations regarding the program. Arriving at appropriate standards is considered a follow-up activity that is conducted once a set of valid criteria is selected.

The broad and inclusive list of potential evaluation criteria identified in Step 2 then provides the base from which the instrument is developed.

Model Step 3: Develop Survey Instrument(s)

The purpose of Step 3 is to meld the potential criteria into a formal survey instrument(s) which representative stakeholders will use to rate the validity and criticality of the potential criteria. Step 3 is comprised of four tasks.

Task 3a. Select scaling components -- no more than two rating scales are suggested. The first rating is for validity and the second for priority.

Scale anchors may vary. Anchors for validity may be a three-point scale consisting of (1) valid, (2) partly valid and partly invalid, and (3) invalid (Guba & Lincoln, 1981). Anchors for priority may be a five-point scale of (5) highest priority to (1) lowest priority.

Task 3b. Select data collection methodology -- Selection of a methodology is based upon such determinants as the number and complexity of the potential criteria, context requirements, population demographics and available resources. Possible alternatives include (1) printed survey sent by mail, (2) printed survey completed from telephone contact, (3) electronic copy sent via computer network or (4) single group decision making session in person, by telephone or by computer.

Task 3c. Prepare survey instrument(s) -- In task 3c, the evaluator prepares the initial instrument(s) using the potential criteria identified in Step 2. The instrument is arranged in a matrix format that permits two ratings for each potential criterion. A section for demographic data

is included at the end of the instrument in order that categorization by stakeholder group can be achieved. A coding scheme may be used to monitor non-responses. A hypothetical example of a survey instrument is attached.

Once the survey instrument(s) are developed, a field-test is conducted using a sample of individuals from the pertinent stakeholder groups. A field-test sampling strategy is chosen which provides an acceptable degree of reliability. After refinements are made, the final versions of the instrument(s) are prepared.

Model Step 4: Collect Data

The purpose of Step 4 is to obtain data that have a high degree of validity and reliability. Step 4 involves the collection of the quantitative data from the survey instrument. There are three tasks in Step 4.

Task 4a. Select sampling strategy -- Selection of a sampling strategy is an evaluator decision based upon the circumstances of the setting. Some settings may suggest a proportionate stratified sample.

Task 4b. Prepare enclosures or attachments -- An introductory letter is prepared by the evaluator which includes a preferred return date. If mailing, a return envelope and postage may be included.

Task 4c. Distribute and follow-up -- The evaluator performs actions to collect the data. If a large sample is used, this may involve a mass mailing or use of a large computer network. If a small, representative sample is used, the data collection may involve telephone calls. When the requested time frame has expired, non-respondents are contacted. If mailing, the non-respondents may be telephoned and a telephone survey will be requested.

Model Step 5: Analyze Data

The purpose of Step 5 is to obtain findings that are used later for selection of the evaluation criterion set. Comparisons within and across stakeholder groups are made as evidence of perceived validity. Step 5 is comprised of three tasks.

Task 5a. Establish cut-off point -- A cut-off point is the point below which criteria are not considered of sufficient practical importance for inclusion. This decision is made by the evaluator and informant group prior to data analysis. If the findings are more extreme than anticipated, this decision may be reconsidered by the evaluator and informant group.

Task 5b. Obtain overall statistics -- Activity 5b involves using an appropriate computer program for obtaining the overall statistics. Means and standard deviations are the primary descriptive statistics used in this model. Data are sorted by stakeholder group to isolate patterns of preference both within and across stakeholder groups.

Task 5c. Prepare matrix of survey findings -- Comparisons within and across stakeholder groups is the final activity in this step. A matrix is developed which reports the findings. This matrix then serves as the guide for selection of the criteria set. A hypothetical example of a matrix of findings is attached.

Model Step 6: Select Criteria Set

The purpose of Step 6 is to establish the validity of a set of program evaluation criteria for a specified mix of stakeholders. Step 6 is comprised of three tasks.

Task 6a. Determine selection scheme -- The evaluator and informant group determine a selection scheme. This involves reviewing the obtained composite scores and making choices; such as choosing a set that includes only the criteria rated high by all stakeholder groups, or choosing a set that includes the top rated criteria of each stakeholder group. The former method is

the method of choice when using a job/task analysis for determining the content of a licensing or certification examination. In both methods, all stakeholder groups are represented although the criteria set may vary between the two selection schemes. Task 6a and 6b may be concluded at a single meeting.

Task 6b. Make final criteria selection -- This may be accomplished by the evaluator alone or with the stakeholder informants.

Task 6c. Document criteria selection -- Once a set of criterion is selected, documentation is prepared to affirm which stakeholder groups are represented.

Summary

The Evaluation Criteria Validation Model is a front-end activity that provides a means for involving stakeholders in selecting a set of evaluation criteria that they perceive as valid for the specified program evaluation. Further research is needed to identify more specifically the issues and concerns involved in using the proposed process.

References:

Byham, William C. and Mortin, Edward Spitzer (1971). *The Law and Personnel Testing*. New York, NY: American Management Association.

Greene, Jennifer C. (1985). *Stakeholder Participation in Evaluation Design: Is It Worth the Trouble?* Paper presented at the Joint Meeting of the Evaluation Research Society, Toronto, Canada, October 1985.

Guba, Egon G. and Lincoln, Yvonna S. (1981). *Effective Evaluation*. San Francisco, CA: Jossey-Bass Publishers.

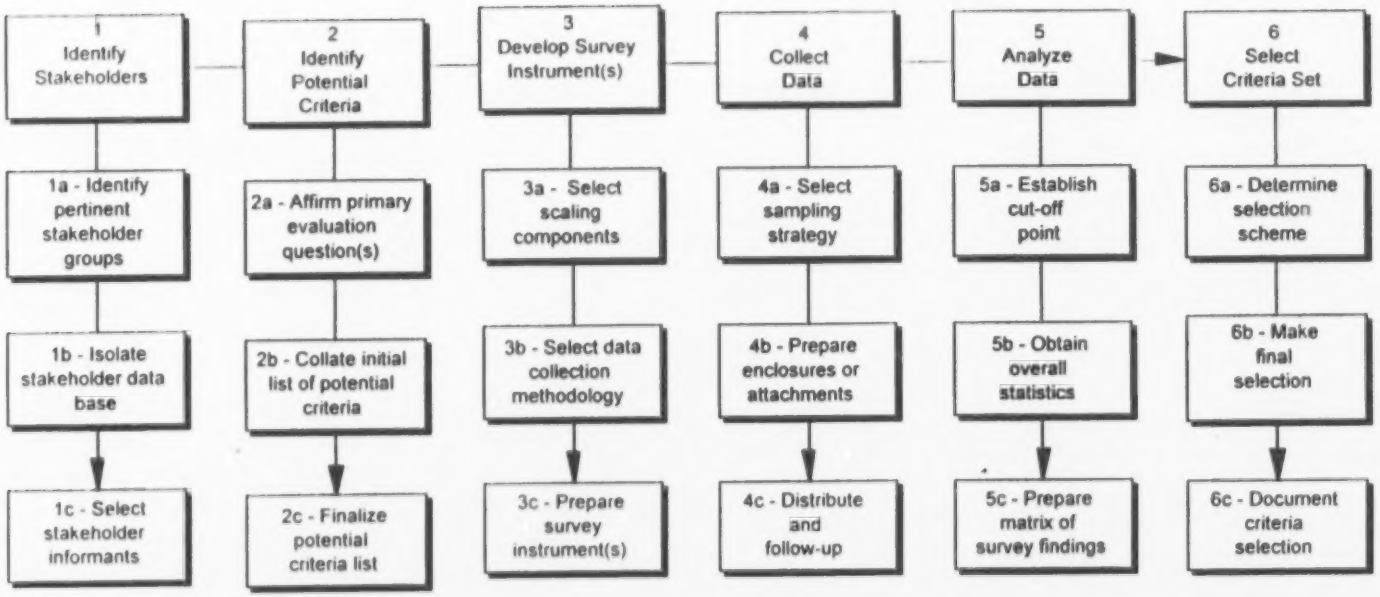
Kane, Michael T.; Kingsbury, Carole; Colton, Dean; and Estes, Carmen, (1989). Combining Data on Criticality and Frequency in Developing Test Plans for Licensure and Certification Examinations. *Journal of Educational Measurement*. Vol.26, pp. 17-27.

Scriven, Michael (1974). The Concept of Evaluation. In *Educational Evaluation: Analysis and Responsibility*. Ed: Apple, Michael W.; Subkoviak, Michael J.; and Lufler, Henry S. Jr. Berkely, CA: McCutchan Publishing Company.

Stake, R. E., (1975). *Evaluating the Arts in Education: A Responsive Approach*. Columbus, OH: Charles E. Merrill.

Thompson, Duane E. and Thompson, Toni A. (1982). Court Standards for Job Analysis in Test Validation. *Personnel Psychology*, Vol.35, pp 865-873.

PROGRAM EVALUATION CRITERIA VALIDATION MODEL


1
Identify Stakeholders
1a - Identify pertinent stakeholder groups
1b - Isolate stakeholder data base
1c - Select stakeholder informants
2
Identify Potential Criteria
2a - Affirm primary evaluation question(s)
2b - Collate initial list of potential criteria
2c - Finalize potential criteria list
3
Develop Survey Instrument(s)
3a - Select scaling components
3b - Select data collection methodology
3c - Prepare survey instrument(s)
4
Collect Data
4a - Select sampling strategy
4b - Prepare enclosures or attachments
4c - Distribute and follow-up
5
Analyze Data
5a - Establish cut-off point
5b - Obtain overall statistics
5c - Prepare matrix of survey findings
6
Select Criteria Set
6a - Determine selection scheme
6b - Make final selection
6c - Document criteria selection

BEST COPY AVAILABLE

HYPOTHETICAL EXAMPLE
MATRIX OF SURVEY FINDINGS
for validating program evaluation criteria set

HYPOTHETICAL STAKEHOLDER GROUPS for professional development course where adult employees and employers select the educational program:

M = Management S = Students/employees
F = Faculty E = Employers

HYPOTHETICAL EVALUATION QUESTION:

What is the worth of this educational program?

HYPOTHETICAL SURVEY INSTRUCTIONS:

In the *shaded* area, rate each potential evaluation criteria for how **valid** you consider each criteria for **judging the worth of this educational program** using this scale:

2 = Valid
1 = Partly Valid and Partly Invalid
0 = Invalid

In the *unshaded* area, rate each potential evaluation criteria for the absolute **priority** you place on each criteria for **judging the worth of this educational program** using a scale ranging from:

5 = Highest priority, to
1 = Lowest priority

	Potential Program Evaluation Criteria	VALIDITY					PRIORITY				
		M	F	S	E	$\bar{x}$	M	F	S	E	$\bar{x}$
1	Objectives - degree objectives are met*	-	-	-	-	-	-	-	-	-	-
2	Objectives - course teaching objectives	2	1	0	0	0.75	5	3	1	1	2.50
3	Objectives - student learning objectives	1	2	2	2	1.75	3	5	4	4	4.00
4	Instructor - experience in the field	2	2	2	2	2.00	4	4	5	4	4.25
5	Instructor - academic credentials	2	2	1	1	1.50	4	4	3	2	3.25
6	Scores - group score on national test	2	2	0	2	1.50	4	4	1	5	3.50
7	Scores - individual score on national test	0	1	2	2	1.25	1	2	4	4	2.75
8	Financial - profitability of the course	2	2	0	0	1.00	5	5	1	1	3.00
9	Financial - fees for taking the course	1	0	2	2	1.25	4	2	5	5	4.00
10	Financial - increased job pay scale	0	1	2	1	1.00	1	2	5	3	2.75

*Insufficiently specific criteria - which type of objectives?

a:\reslsamp.tab (wp)

HYPOTHETICAL EXAMPLE
SURVEY INSTRUMENT
for validating program evaluation criteria set

HYPOTHETICAL STAKEHOLDER GROUPS for professional development course where adult employees and employers select the educational program:

M = Management S = Students/employees
F = Faculty E = Employers

HYPOTHETICAL EVALUATION QUESTION:

What is the worth of this educational program?

HYPOTHETICAL SURVEY INSTRUCTIONS:

In the *shaded* area, rate each potential evaluation criteria for how **valid** you consider each criteria for **judging the worth of this educational program** using this scale:

2 = Valid
1 = Partly Valid and Partly Invalid
0 = Invalid

In the *unshaded* area, rate each potential evaluation criteria for the absolute **priority** you place on each criteria for **judging the worth of this educational program** using a scale ranging from:

5 = Highest priority, to
1 = Lowest priority

	Potential Program Evaluation Criteria	VALIDITY			PRIORITY				
		2	1	0	5	4	3	2	1
1	Objectives - degree objectives are met*								
2	Objectives - course teaching objectives								
3	Objectives - student learning objectives								
4	Instructor - experience in the field								
5	Instructor - academic credentials								
6	Scores - group score on national test								
7	Scores - individual score on national test								
8	Financial - profitability of the course								
9	Financial - fees for taking the course								
10	Financial - increased job pay scale								

*Insufficiently specific criteria - which type of objectives?

a:\critsamp.tab (wp)